FlyBase: anatomical data, images and queries

Gary Grumbling*, Victor Strelets and The FlyBase Consortium

Department of Biology, Indiana University, 1001 E 3rd Street, Bloomington, IN 47405-7005, USA

ABSTRACT

FlyBase (http://flybase.org/) is a database of genetic and genomic data on the model organism *Drosophila melanogaster* and the entire insect family Drosophilidae. The FlyBase Consortium curates, annotates, integrates and maintains a wide variety of data within this domain. Access to the data is provided through graphical and textual user interfaces tailored to particular types of data. FlyBase data types include maps at the cytological, genetic and sequence levels, genes and alleles including their products, functions, expression patterns, mutant phenotypes and genetic interactions as well as aberrant chromosomes, annotated genomes, genetic stock collections, transposons, transgene constructs and insertions, anatomy and images, bibliographic data, and community contact information.

AN OVERVIEW OF FlyBase

FlyBase is the model organism database for *Drosophila melanogaster*. It is the principal resource for curated and integrated genetic and molecular data for the insect family Drosophilidae (1). Over this broad scope, data are captured and combined by curators and annotators with attribution to a primary source being an underlying principle. The types of data made available for these species is diverse and can be accessed through a variety of interfaces and tools generally organized into data classes on the home page (http://flybase. org/; see Supplementary Figure 1).

FlyBase includes genetic, cytogenetic and annotated genomic maps that can be browsed or queried by location, gene or any of several other criteria. Genome annotations for *D.melanogaster* [reviewed in (2)] and *Drosophila pseudoobscura* (3), currently at versions 4.2 and 2.0 respectively, are accessible through the Generic Genome Browser (GBrowse) (4) (http://flybase.org/cgi-bin/gbrowse_fb/dmel; see Supplementary Figure 2) as well as through annotation reports for individual genes. BLAST and sequence download services are also provided for these and other genomes. FlyBase provides access to information in a range of other categories including phenotypes (5), genetic interactions, chromosomal aberrations, transgene constructs and insertions, natural transposons and transposon insertions, anatomical data and associated images, experimental resources such as stocks of genetic variants and genomic clones, extensive bibliographic data, and information about researchers in the community. Table 1 gives an overview of selected FlyBase content as on September 2005.

Notable recent changes at FlyBase include the complete integration of the *D.pseudoobscura* genome and availability of preview sequence data for other *Drosophila* genomes. Access to these data is being provided in advance of GenBank submission as a service of FlyBase together with the genome sequencing centers. The FlyBase BLAST service has been updated as a result and now accommodates multiple species (http://flybase.org/blast/; see Supplementary Figure 3). New stock lists have been added, including the collection of the Tucson *Drosophila* Species Stock Center (http://stockcenter. arl.arizona.edu/), as have new linkouts to external resources. FlyBase now provides computed orthology calls to other organisms based on data provided by the InParanoid project (6) (http://inparanoid.cgb.ki.se/index.html). The 'Anatomy and Images' section of FlyBase has also undergone extensive renovation. The remainder of this article focuses on anatomical data, associated images and their access in FlyBase.

ANATOMICAL DATA IN FlyBase

Many FlyBase data types have in common their description with a structured controlled vocabulary (CV) or ontology. The use of CVs allows both for consistent curation practices and improved search and retrieval for users. For example, FlyBase uses the Gene Ontology (GO) (7), the Cell Type Ontology (8)

*To whom correspondence should be addressed. Tel: +1 812 856 2381; Fax: +1 812 855 6705; Email: ggrumbli@bio.indiana.edu

Table 1. Categories of data in FlyBase and their current contents as on September 2005

Drosophila melanogaster genes	~28 300
Non-*melanogaster* Drosophilid genes	~23 000
Protein-coding and non-coding RNA genes (all species)	~48 000
D.melanogaster genes with genome annotations	~14 400
Drosophila pseudoobscura genes anchored to the genome with *D.melanogaster* orthologs	~12 200
D.pseudoobscura genes with genome annotations	~10 000
Genes with GO annotations	~10 700
Natural transposons	~1100
Insertions of natural transposons in *D.melanogaster* sequenced strain	~6000
Genetically engineered transposons (transgene constructs)	~18 900
Transgene insertions	~55 500
Stock center mutant and wild-type strains (all species)	~28 900
D.melanogaster mutant alleles	~71 600
Non-*melanogaster* Drosophilid mutant alleles	~5100
Phenotypic data CV statements	~162 200
Genetic interaction CV statements	~70 500
Unique anatomy CV terms associated with phenotypes	~2300
Unique anatomy CV terms associated with gene expression patterns	~1000
Unique anatomy CV terms annotated to images	~1000
Anatomical images with annotations	~1000

and the Sequence Ontology (SO) (9) for describing aspects of genetic and molecular data. It also makes extensive use of CVs that describe the Anatomy and Development of the fly in both spatial and temporal terms. These Anatomical and Developmental vocabularies were developed over time within FlyBase and are now available as part of the Open Biomedical Ontologies (OBO) project (http://obo.sourceforge.net/).

The anatomy and development of the fly is one very important aspect of the dataset curated into FlyBase, and the use of these ontologies is essential to the curation of those data. The terms from the fly anatomical and developmental ontologies are used in the description of mutant phenotypes and genetic interactions and also in the description of transcript and protein expression patterns. Gene subreports available from a 'Synopsis' gene report (Figure 1) include anatomical and developmental data. The 'Alleles' subreport has these data in the 'Phenotype manifest in' field under the 'Summary of Allele Phenotypes' section. This field shows all the CV terms that describe a given allele with attribution to a particular reference from which the data were curated. The 'Genetic Interactions' subreport has anatomy and development data under the 'Summary of Genetic Interactions' section in the 'Genetic

Figure 1. The FlyBase 'Synopsis' gene report (http://flybase.org/.bin/fbidq.html?FBgn0003984&content=onepage). Following the links in the highlighted subreports and subsection leads to reports that contain fly anatomy and development CV terms used in the curation of phenotype and expression pattern data (see also Figure 2).

Expression pattern

Stage	Tissue / Position	Pattern	Source
embryo	ventral midline, epidermis, central nervous system, labral segment, maxillary lobe, labial segment, stomodeum	--	vn-XR
embryo | blastoderm	--	stripes | ventrolateral	vn-XR
embryo | blastoderm	amnioserosa	--	vn-XR
embryo | stage 10	tracheal pit	--	vn-XR
embryo | stage 16	somatic muscle, visceral muscle	--	vn-XR
embryo | stage <=11	amnioserosa	--	vn-XR
embryo | stage >=11	peripheral nervous system, Keilin's organ, chordotonal organ	--	vn-XR
larva	eye-antennal disc, dorsal metathoracic disc, ventral thoracic disc	--	vn-XR
larva, pupa	dorsal mesothoracic disc, scutum	--	vn-XR
pupa	wing cell	--	vn-XR

Summary of Allele Phenotypes

Phenotype manifest in	Allele
abdominal 2 dorsal acute muscle 1, abdominal 2 lateral longitudinal muscle 1, abdominal 2 ventral acute muscle 2, abdominal 3 dorsal acute muscle 1, abdominal 3 lateral longitudinal muscle 1, abdominal 3 ventral acute muscle 2, abdominal 4 dorsal acute muscle 1, abdominal 4 lateral longitudinal muscle 1, abdominal 4 ventral acute muscle 2, abdominal 5 dorsal acute muscle 1, abdominal 5 lateral longitudinal muscle 1, abdominal 5 ventral acute muscle 2, abdominal 6 dorsal acute muscle 1, abdominal 6 lateral longitudinal muscle 1, abdominal 6 ventral acute muscle 2, abdominal 7 dorsal acute muscle 1, abdominal 7 lateral longitudinal muscle 1, abdominal 7 ventral acute muscle 2	vn^{L6}
abdominal 2 dorsal acute muscle 1, abdominal 3 dorsal acute muscle 1, abdominal 4 dorsal acute muscle 1, abdominal 5 dorsal acute muscle 1, abdominal 6 dorsal acute muscle 1, abdominal 7 dorsal acute muscle 1, embryonic/larval somatic muscle	vn^{10567}
adult cuticle, head capsule, ocellar bristle, ocellus, postvertical bristle	vn^{ddd-3}
anterior crossvein	vn^{C221}
anterior crossvein, posterior crossvein, wing vein L3, wing vein L4	vn^{1}
anterior crossvein, wing	vn^{ddd-13}

Figure 2. The 'Expression pattern' and 'Summary of Allele Phenotypes' sections of the 'Expression and Phenotypes' subreport (http://flybase.org/.bin/fbidq. html?FBgn0003984&content=phenotype). The linked terms in the 'Tissue/Position' fields describe the expression pattern of the transcript in the 'Source' column at particular points in time and lead to 'CV term' reports. Similarly, the linked terms in the 'Phenotype manifest in' fields describe the phenotype of the particular allele in the 'Allele' column and lead to 'CV term' reports. The terms in this example show the wide variety of fly biology that can be described using the Anatomy and Development ontologies.

interaction' anatomy fields. The 'Proteins and Transcripts' subreport has these data under the 'Expression pattern' section. Here expression pattern data for gene products are described in terms of stage, tissue or position and pattern fields. Each expression pattern statement is also attributed to a particular reference.

Anatomical and developmental data can also be found in a 'Synopsis' gene report in abbreviated form in the 'Expression and Phenotypes' subsection. The terms listed next to the 'Expressed in' and 'Mutants affect' labels give a brief description of the mutant phenotypes and expression patterns that have been curated for a gene. The link labeled 'details . . .' leads to complete information available for a given gene in the 'Expression and Phenotypes' subreport (Figure 2). The 'Expression and Phenotypes' subreport is the combination of other presentations that puts in one report both phenotype and expression pattern data for a given gene.

THE VOCABULARY TERM AND ANATOMICAL IMAGE REPORTS

There are two interrelated reports provided by FlyBase that are particularly focused on anatomy and development, the 'CV term' report and the 'Image' report. The anatomical data outlined above is used as an organizing principle to unify many types of data in FlyBase in the 'CV term' report. The user is able to gain an anatomical perspective on subsets of the data by drawing together data in FlyBase that have in common annotation to a certain CV term. FlyBase currently makes available reports for CV terms from the GO, the Cell Type Ontology and the fly Anatomy and Development ontologies.

These terms can be accessed through a search tool called 'TermLink', which allows for both searching and browsing of the ontologies (http://flybase.org/cgi-bin/fbcvq.html?start). This tool supports access to FlyBase 'CV term' reports for anatomical terms, such as 'arista' (Figure 3). This report shows the fields that describe a term and their values, including a definition and any synonyms for the term. The location of the term in the hierarchy and its relationship to parent and child terms is shown in order to put the term in context. The levels of the term hierarchy displayed can be controlled through options at the bottom of the page. Links to other objects in FlyBase that have also been associated with this term are provided. These associated objects include genes, alleles, polypeptides, transcripts and images. Following the links leads to query results for the objects associated with this term and then to the specialized reports for the chosen type of object.

FlyBase CV term report

Term	arista
Definition	The strong bristle on the 3rd antennal joint of some Diptera, that may be simple, jointed, glabrous or ciliate.
ID	**FBbt:00004519**
Synonyms	segment 6
Relationship	part_of FBbt:00004512 ! adult antennal structure

Parents/Children (Spanning Tree)	antenna
	  |__adult antennal structure__
	antennal segment |
	  |__adult antennal structure__|
	                                    arista
	                                      |__arista central core
	                                      |__arista lateral
	                                      |__aristal nerve
	                                      |__aristal sense organ__

Associated FlyBase objects	Search for Genes Alleles Transcripts Polypeptides Images
Spanning Tree View Settings	Show hierarchy levels: 2 ▾ for parents, 1 ▾ for children Re-Draw

Figure 3. The FlyBase 'CV term' report for the CV term 'arista' (http://flybase.org/.bin/cvreport.html?arista). This report shows fields that describe the term, displays the location of the term in a subset of the term hierarchy and has links to other objects in FlyBase that are associated with this term.

For example, following the link to associated images from the 'CV term' report for 'arista' opens a page of 'Images query results' for this term listing seven thumbnail preview images returned for this query. These consist of graphic representations of the head and antennal structures. Choosing the image with a short description of 'The mesal aspect of the left antenna' leads to an 'Image' report that includes this structure.

The 'Image' report has fields describing the anatomical image and links to associated anatomical terms (Figure 4). The 'Image' report also supplies attribution for, and a link to, the source from which the image was obtained. Anatomical images are acquired from the literature and other sources where copyright permission can be obtained. The images are annotated with terms from various CVs to support their retrieval. Finally, images have maps of areas corresponding to each CV term overlaid, allowing the relevant anatomical structures to link to the anatomy terms used to annotate the image. When a user rolls the cursor over an area of anatomy or its corresponding term both the area and term are highlighted (shown in Figure 4 for the arista). This feature provides an online, interactive replacement for the original figure legend that is normalized to the vocabulary used by FlyBase. Following the link on a term or a mapped area returns one to the 'CV term' report.

The 'Anatomy and Images' section of FlyBase supports the selection of images for browsing, and provides alternative entries into the anatomical and developmental vocabularies based on the concepts of developmental stage or life cycle, tagma or body segment, organ system, germ layer and species (http://flybase.org/anatomy/). The images in this section of FlyBase that have been acquired, described and united with terms from the ontologies represent and illustrate the biology of the fly. Accordingly, the images serve both to illustrate the anatomical and developmental vocabularies and to provide users with some orientation within these large and complex vocabularies. For example, the fly anatomy ontology
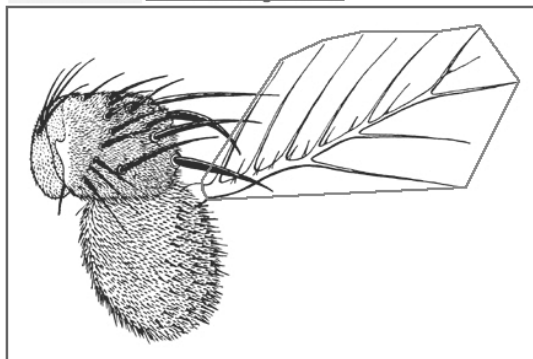
encompassed 6125 terms at the time of this writing. This is important because these vocabularies are central to how such data in FlyBase are captured and organized. The FlyBase image collection also functions as a tutorial aid on the anatomy and development of the fly that connects to the appropriate references. There are other online resources that address similar needs, some of which are listed in the 'Atlases and images' section of the FlyBase '*Drosophila* Resources' page (http://flybase.org/allied-data/resources.html#atlas). The aim of the 'Anatomy and Images' section of FlyBase is to make available images to graphically represent as many as possible of the anatomical and developmental terms used by FlyBase. To this end, we will, in the future, be adding more images to the collection, with more specialized high level entry points into these data.

FUTURE DIRECTIONS FOR FlyBase

The FlyBase project will see changes as well. The sequence and analysis of ten more species genomes from the *Drosophila* comparative genomics sequencing projects will be incorporated into FlyBase in the coming months when the annotated genomes are deposited in GenBank. Announcements regarding the status of these genomes will be posted on the FlyBase home page in the 'Important News' area. FlyBase will provide the computed annotations and access to them through GBrowse as it now does with *D.pseudoobscura* (see Supplementary Figure 2) and will supply gene reports for the genes in these species. BLAST access to the sequences (see Supplementary Figure 3) will continue to be provided as it is now, ahead of the availability of assemblies. There will be further annotation releases of the *D.melanogaster* and *D.pseudoobscura* genomes with the *melanogaster* sequence version 5.0 expected to be available soon. FlyBase is in the process of migrating its data to the GMOD Modular Schema,

FlyBase Image report

Name	Ferris, 1950 Fig 03 A
FlyBase ID	FBim1662283
Description	The mesal aspect of the left antenna. After Figure 3 A in Ferris, G.F. 1950. External morphology of the adult. Demerec, 1950 368--419. In Demerec, M. 1994. Biology of Drosophila. [Facsimile edition.] Copyright 1994 by Cold Spring Harbor Laboratory Press. Used with kind permission from Cold Spring Harbor Laboratory (CSHL) Press, see http://www.cshl.org/.
Date	30-Aug-02
Reference	FBrf0007734
Categories	adult stage adult head integumentary system
Anatomy Terms	arista antennal segment 5 antennal segment 3 antennal segment 2 antennal segment 1

Figure 4. The anatomical 'Image' report for an image that graphically represents the arista (http://flybase.org/.bin/imagereport.html?FBim1662283). Fields in this report describe the image and show all the terms used to annotate the image. If subcomponents of the image have been demarcated, the area of the image that corresponds to a given term and the term in the 'Anatomy Terms' field will simultaneously be highlighted when the cursor rolls over either of them. This creates the equivalent of a figure legend with annotation that has been normalized to the CV used by FlyBase.

or Chado (http://www.gmod.org/schema/), which is one component of the Generic Model Organism Database (GMOD) project (http://www.gmod.org/). As this process moves forward, we plan an overhaul and redesign of our current web user interface in the coming year based on a fully integrated dataset being available in the new schema.

REFERENCING FlyBase

We suggest FlyBase be referenced in publications by citing this publication and the FlyBase web address (http://flybase.org).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Drysdale,R., Crosby,M. and FlyBase Consortium (2005) FlyBase: genes and gene models. *Nucleic Acids Res.*, **33**, D390–D395.
2. Misra,S., Crosby,M., Mungall,C., Matthews,B., Campbell,K., Hradecky,P., Huang,Y., Kaminker,J., Millburn,G., Prochnik,S. *et al.* (2002) Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol.*, **3**, RESEARCH0083.
3. Richards,S., Liu,Y., Bettencourt,B., Hradecky,P., Letovsky,S., Nielsen,R., Thornton,K., Hubisz,M., Chen,R., Meisel,R. *et al.* (2005) Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and *cis*-element evolution. *Genome Res.*, **15**, 1–18.
4. Stein,L., Mungall,C., Shu,S., Caudy,M., Mangone,M., Day,A., Nickerson,E., Stajich,J., Harris,T., Arva,A. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.
5. Drysdale,R. (2001) Phenotypic data in FlyBase. *Brief. Bioinform.*, **2**, 68–80.
6. O'Brien,K., Remm,M. and Sonnhammer,E. (2005) Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.*, **33**, D476–D480.
7. Harris,M., Clark,J., Ireland,A., Lomax,J., Ashburner,M., Foulger,R., Eilbeck,K., Lewis,S., Marshall,B., Mungall,C. *et al.* (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.*, **32**, D258–D261.
8. Bard,J., Rhee,S. and Ashburner,M. (2005) An ontology for cell types. *Genome Biol.*, **6**, R21.
9. Eilbeck,K., Lewis,S., Mungall,C., Yandell,M., Stein,L., Durbin,R. and Ashburner,M. (2005) The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.*, **6**, R44.